HHS Public Access
Author manuscript
J Res Educ Eff. Author manuscript; available in PMC 2018 July 09.

Published in final edited form as:
J Res Educ Eff. 2018 ; 11(3): 339–374. doi:10.1080/19345747.2018.1441347.

Do High-Quality Kindergarten and First-Grade Classrooms Mitigate Preschool Fadeout?

Jade M. Jenkins[1,*], **Tyler W. Watts**[1], **Katherine Magnuson**[2], **Elizabeth Gershoff**[3], **Douglas Clements**[4], **Julie Sarama**[4], and **Greg J. Duncan**[1]

[1]School of Education, University of California, Irvine

[2]School of Social Work, University of Wisconsin, Madison

[3]Department of Human Development and Family Sciences, University of Texas at Austin

[4]College of Education, University of Denver

Abstract

Prior research shows that short-term effects from preschool may disappear, but little research has considered which environmental conditions might sustain academic advantages from preschool into elementary school. Using secondary data from two preschool experiments, we investigate whether features of elementary schools, particularly advanced content and high-quality instruction in kindergarten and first grade, as well as professional supports to coordinate curricular instruction, reduce fadeout. Across both studies, our measures of instruction did not moderate fadeout. However, results indicated that targeted teacher professional supports substantially mitigated fadeout between kindergarten and first grade but that this was not mediated through classroom quality. Future research should investigate the specific mechanisms through which aligned preschool-elementary school curricular approaches can sustain the benefits of preschool programs for low-income children.

Keywords

Head Start; Preschool; Fadeout; Early Childhood Education; Classroom Quality

Introduction

Early childhood education (ECE) experiences improve children's school readiness, with low-income and disadvantaged children appearing to benefit the most from these programs (Barnett, 2011; Camilli, Vargas, Ryan, & Barnett, 2010; Duncan & Magnuson, 2013; Reynolds, Temple, & Ou, 2010). Preschools, ECE for three- to-five-year-olds, often use curricula to guide their classroom learning activities. Curricula vary in how they function, but nearly all set at least broad learning goals and provide a set of suggested day-to-day activities and experiences intended to help children reach the goals (Goffin & Wilson, 1994; Ritchie & Willer, 2008). Research suggests that not all curriculum are equally effective at

*Corresponding author: jvjenkin@uci.edu; 3200 Education, Irvine, CA 92617; 949-824-7987.

boosting children's early skills; some preschool curricula generate significantly more learning gains when compared with "business as usual" preschool classroom activities.

For early childhood education programs and their curricula, many evaluation studies find that children's early learning gains found at program completion do not persist through the elementary school years. Often, after a year or two of schooling differences between preschoolers and comparison groups have almost entirely disappeared (Barnett, 1995; Bassok, Gibbs, & Latham, 2015; Currie, 2001; Puma, Bell, Cook, & Heid, 2010). After analyzing existing preschool studies, both Aos and colleagues (2006) and Li and colleagues (2016) found that the end-of-treatment preschool impacts on cognitive and achievement outcomes decline by half in the year following end of preschool, and then by half again over the next two years. Although this decline in program impacts overtime is not unique to early childhood interventions, it is discouraging to policymakers and practitioners because promoting later school achievement is one of the core motivations for funding public early education programs (Bailey, Duncan, Odgers, & Yu, 2017).

The convergence in academic and cognitive skills between treatment and control-group children in early elementary school is often termed preschool "fadeout" or control group "catch-up," and they are two sides of the same phenomenon. A preschool impact only persists as long as the children who attended preschool continue to learn new material in elementary schools at the same or a faster rate than the children who have not had any early childhood education learn new material. As new material becomes increasingly complex in the early school years and requires additional cognitive skills and efforts, this advantage may be hard to sustain.

Although common in the evaluation literature, preschool fadeout in the early school years has received little theoretical and empirical attention. One longstanding explanation of fadeout is that low-income preschool graduates enter schools that do not support their prior academic gains (Brooks-Gunn, Markman-Pithers, & Rouse, 2016; Currie & Thomas, 2000; Lee & Loeb, 1995). Studies of kindergarten classrooms show that a great deal of instruction covers content that students already know (Engel, Claessens, & Finch, 2012; Gervasoni & Perry, 2015) – a mismatch that can possibly be magnified by preschool learning. Moreover, disadvantaged children may be more likely to attend schools where teachers are under intense pressure to meet proficiency standards and do not differentiate instruction or teach advanced content to students who already know the basics (Darling-Hammond, 2004; Stipek, 2004). If the classroom instructional practice is better matched to the learning needs of children who have not had enriching early learning experiences, then those that attend preschool may learn less than those who did not.

A related, but broader concern, is that low-quality schools do not serve the instructional needs of any children well. As a result all children's skills stagnate at a low level and the program impacts would not last long. Fewer resources, poorly-managed classrooms, and the distractions of dangerous conditions are more prevalent in elementary schools attended by students in low-income communities (Lee & Loeb, 1995) and each could hinder preschoolers' ability to sustain learning gains.

Even if early instruction could be improved for preschool attendees by differentiated instruction or other means, the net impact on children entering school with and without preschool experience is ambiguous. Yes, enriched, higher-quality classrooms may well enhance the early-grade learning of preschool graduates, but it could also boost the learning of their control-group counterparts even more. Correlational evidence shows that all children benefit from exposure to advanced reading and math content, regardless of whether they attended preschool, began school with stronger skills, or are from families with low income (Engel et al., 2012). Thus, even if children "treated" in their preschool year gain more from higher- than lower-quality early-grade classrooms, fadeout may still occur as long as classroom quality effects are even larger for children who did not attend preschool.

These issues motivate our study of whether preschool graduates' instructional experiences in elementary school moderate the persistence of preschool's effects. We used two studies of preschool interventions that included follow-up data on children's elementary school environments: Head Start and the TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development) scale-up intervention of a mathematics curriculum. The Head Start study allows us to examine whether the rigor of kindergarten and first-grade instructional content influences the long-run impact of participation in Head Start. Because the TRIAD study randomly assigned its preschool curriculum treatment group children to a follow-up pedagogical support intervention in kindergarten and first grade, it provides an additional opportunity to examine how later instructional experiences may affect the persistence of preschool program impacts. Whereas the Head Start study compares attending or not attending preschool on sustaining impacts, the TRIAD study is a contrast between attending preschool with a supplemental instructional intervention and preschool without a focused math curriculum. Including both of these interventions in our study enables us to conduct a robust examination of whether the instructional features of children's elementary school environments matter for sustaining the effects of both types of preschool interventions, and provides a built-in replication study of long-standing hypotheses about preschool fadeout and persistence.

Background

Fadeout in the effects of public preschool

The U.S. has seen a rapid expansion of ECE programs over the past 40 years, an expansion made possible in part because policymakers and educators now view early childhood as a particularly opportune time for investment (Heckman, 2006; Jenkins, 2014). Such views have been shaped by long-run experimental evidence from some model programs (i.e., Perry Preschool, Abecedarian), showing that, in addition to end-of-preschool impacts on school readiness, preschool participants are less likely to be retained or drop out of high school, as well as more likely to attend a 4-year college compared with children who did not attend preschool, and have increased rates of employment and earnings, as well as lower rates of adult poverty and arrest, perhaps as a result of increased educational attainment and skill (Barnett & Masse, 2007; Belfield, Nores, Barnett, & Schweinhart, 2006; Campbell et al., 2014; Campbell et al., 2012; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Campbell et al., 2008; Schweinhart, 2005). Strong quasi-experimental studies find substantial intermediate effects of attending state preschool programs on school achievement in the elementary and middle school years (Andrews, Jargowsky, & Kuhne, 2012; Cascio & Schanzenbach, 2013). Notably, a series of studies by Dodge, Ladd, & Muschkin (2017; 2014; 2015) on North Carolina's public preschool funding found positive impacts on achievement, and a reduction in grade retention, and special education placement through the end of fifth grade. Finally, rigorous quasi-experimental studies on Head Start participants find long-run positive impacts on academic and health outcomes of .2-.3 standard deviations (SD) in adulthood (Currie & Thomas, 1995; Deming, 2009; Garces, Thomas, & Currie, 2002; Ludwig & Phillips, 2008).

Yet the larger literature of ECE studies provides a more complicated pattern of end-of-preschool (short-term) effects. Most studies show evidence of initial impacts on cognitive or early academic outcomes, but also suggest that these impacts diminish significantly in 1-3 years after preschool as comparison children catch up to preschool attendees. Program impact fadeout of achievement and cognitive outcomes has been especially noted in more recent studies of large public programs including Head Start and some state preschool programs (Hill, Gormley, & Adelstein, 2015; Lipsey, Farran, & Hofer, 2015; Puma et al., 2012).

Perhaps the best-known instance of preschool fadeout was observed in the Head Start Impact Study experiment (HSIS). This was the first nationally representative random assignment evaluation of the federal preschool program for low-income children. In 2002, two cohorts of children were randomly assigned to receive Head Start services at sites across the country. The end-of-program-year impact effect sizes averaged .2 SD for both the age-3 and age-4 cohorts on early language and literacy skills, and a .15 SD effect size was observed on early math skills for age-3 cohort participants (Puma et al., 2010). These results were about 50 percent larger when non-compliance with treatment assignment was taken into account (Ludwig & Phillips, 2008). Nonetheless, the modest short-term gains from Head Start were entirely gone by the study's follow-up periods in elementary school (kindergarten, first, and third grade; Puma et al., 2012). Although the quick convergence of control and treatment group test scores was concerning, it aligned with an earlier study finding that Head Start produced long-term impacts on important outcomes such as educational attainment, that rebounded after initial fadeout (Deming, 2009). Questions remain regarding what might account for the discrepancies between findings from earlier small-scale random assignment studies and later larger-scale quasi-experimental research.

To answer such questions, recent studies have tried to examine the mechanisms by which program effects fade or persist. Perry and Abecedarian follow-up studies suggest that substantial proportions of the adult impacts were explained by measures of children's cognitive and socioemotional skills (Elango, García, Heckman, & Hojman, 2015; Heckman, Pinto, & Savelyev, 2013). Because Perry's large impacts on IQ had completely disappeared by age 8, while impacts on an assortment of personality and behavior measures persisted beyond that point, "noncognitive" skills, such as conscientiousness and executive function, are viewed as the primary channel of sustained impacts (Heckman et al., 2013). One of the most comprehensive attempts to understand the processes by which an ECE program affected later outcomes is the Reynolds, Ou, and Topitzes (2004) analysis of data from the

Chicago Child-Parent Centers, which provided, among other benefits, half-day preschool at ages 3 to 4 years and half- or full-day kindergarten to children living in Chicago's Southside. They found long-run effects on completed schooling, and analyses suggested that cognitive test scores served as mediators, but motivational measures (e.g., "I try hard in school") gathered at age 11 did not. However, these studies also find that around half of the adult impacts cannot be explained by measured cognitive or noncognitive skills, leaving questions as to whether environmental factors could play a role in determining fadeout or persistence of early intervention impact. We investigate this possibility, testing whether elementary school classroom instruction plays a role in preschool fadeout.

Beyond "preschool as usual": Fadeout in preschool curricular interventions

In many ECE programs, the longstanding presumption has been that simply exposing children to a diverse and stimulating set of activities will result in learning. However, some scholars (Stipek, 2006) note that few preschool programs design such activities in an optimal way to support early learning. Indeed, if learning activities and instruction are more carefully planned and align with developmental trajectories, then preschool may have a stronger impact on children's learning and the persistence of such impacts. As such, another target of interventions to improve children's school readiness is the content and nature of preschool instruction through the classroom curriculum.

Like preschool, however, the effects of curricular interventions also fade out. Curricula set goals for the knowledge and skills that children should acquire in an educational setting, and support educators' plans for providing the day-to-day learning experiences to cultivate those skills with daily lesson plans, materials, and other pedagogical tools (Goffin & Wilson, 1994; Ritchie & Willer, 2008). Across the U.S. the curricula representing "business as usual" preschool instruction follow a "whole-child" approach (Jenkins & Duncan, 2017). "Whole-child" preschool curricula are broad in content and seek to generate gains in cognitive, physical, social, and emotional domains of children's development (Diamond, 2010; Elkind, 2007; Zigler & Bishop-Josef, 2006). Innovations and interventions in this area involve preschool curricula designed using a narrower, scholastic focus (i.e., target literacy or math skills), with some showing positive impacts on skills targeted in the curricular materials (Bierman et al., 2008; D. H. Clements & Sarama, 2008; Diamond, Barnett, Thomas, & Munro, 2007; Fantuzzo, Gadsden, & McDermott, 2011; Morris et al., 2014). Notably, experimental evidence of the *Building Blocks* preschool mathematics curriculum, which encourages the acquisition of conceptual and procedural knowledge in both numeracy and geometric/spatial reasoning through the emphasis of empirically-supported learning trajectories (see D. H. Clements & Sarama, 2008), shows substantial increases in children's mathematics achievement by the end of preschool (Hedge's g = 0.71; D. H. Clements, Sarama, Spitler, Lange, & Wolfe, 2011). However, this impact had shrunk by nearly 60% by the end of first grade. Few other studies have followed children long enough to provide information about the degree to which curriculum impacts persist or fade over time.

In sum, the literature on fadeout and persistence in preschool impacts is equivocal, leaving policymakers and researchers trying to reconcile fadeout in short-term academic impacts in recent programs with the possibility of long-term benefits for other related outcomes from programs runs decades ago. One common explanation for fadeout focuses on the nature of children's subsequent learning contexts. But much more attention is needed to better understand why and how the persistence of both preschool program and preschool curriculum impacts are affected by later instructional contexts.

Sustaining School Environments

Bailey and colleagues (2017) use the term "sustaining environments" to refer to the idea that the quality of a child's environments subsequent to the completion of a preschool intervention may be important for sustaining early skill advantages. The idea that poor subsequent environments erode preschool gains is premised on the fact that children from low-income families enter quality and resource-poor schools because school finances are tied to local property taxes in most states and localities, and disadvantaged children tend to live in property-poor neighborhoods (M. A. Clements, Reynolds, & Hickey, 2004; Crosnoe & Cooper, 2010; McLoyd, 1998; Pianta, Belsky, Houts, & Morrison, 2007; Stipek, 2004). More numerous school-supporting nonprofit organizations (e.g., PTAs) in wealthier districts with higher average education levels also exacerbate disparities in per-pupil expenditures between high- and low-income students (Nelson & Gazley, 2014).

When children in low-income areas leave preschool and begin kindergarten in resource-poor schools, such schools may be ill-equipped to build upon the skills children gained during preschool (Currie & Thomas, 2000; Lee & Loeb, 1995; Reynolds et al., 2004; Zhai, Raver, & Jones, 2012). Children from low-income families benefit the most from consistently cognitively-stimulating environments (Crosnoe et al., 2010), which suggests that the schools they attend may be missing opportunities to improve the achievement and attainments of these children. Indeed, recent research suggests that the benefits from attending Head Start were largest when children followed Head Start with enrollment in well-funded K-12 schools (Johnson & Jackson, 2017). However, studying resources alone leaves questions regarding which classroom or school processes sustain learning gains. For policies to encourage development across preschool through the K-12 years, essential to our understanding is whether and how instructional processes can reduce preschool fadeout.

Sustaining classroom instruction—Preschool program or curricular impacts will persist if children who received treatment learn *at the same or higher rate* than those who did not attend preschool or receive the curriculum. In the case of resource-poor schools, instructionally-poor early-grade classrooms may enable children not attending preschool to learn basic skills but are unlikely to build on the higher school-entry skills of preschool attendees. In this case, preschool impacts fade out because the poor-quality early-grade classrooms enable non-attendees to learn at a faster rate and catch-up with the attendees. However, few scholars would argue that low-quality instruction is beneficial for children entering school with fewer skills; enriched instructional experiences should enhance learning among all children.

A variation of this argument is that high-quality instruction *differentially* benefits the academic skills of preschool graduates by continuing their skilled growth at the same or faster pace than their non-attending peers (Barnett, 2011; McKey et al., 1985; Swain, Springer, & Hofer, 2015; Zigler & Styfco, 2004). This is similar to the "skills beget skills" hypothesis (Cunha, Heckman, & Schennach, 2010; Heckman, 2006; Miller, Farkas, Vandell, & Duncan, 2014), in that preschool attendees enter kindergarten with more advanced academic skills than their non-attending peers, and are better primed to benefit from good instruction.

A useful way to categorize sustaining classroom instruction is by how skillfully teachers teach (i.e., pedagogy, teacher-child interactions), what content they cover (i.e., topics and difficulty level) and the extent to which instruction meets the needs of different kinds of students (Connor et al., 2009; Early et al., 2010; National Mathematics Advisory Panel, 2008; Yoshikawa et al., 2013). Developmental theory suggests that regular exposure to content that is both beyond a child's current skill level and still within their range of abilities is critical for children's intellectual development (Bronfenbrenner, 1989; Vygotsky, 1978). This means that when preschool attendees enter elementary school with the foundational early skills learned during preschool (e.g., letter recognition, cardinality), they should be exposed to sequentially more challenging tasks and concepts as they progress through the early grades for continued cognitive development and sustained preschool impacts.

Coordinating instructional content between preschool to kindergarten may help to reduce instructional repetition and ensure preschool attendees continue to be challenged academically. Clements and colleagues (2013) demonstrated that aligning early-grade and preschool instruction can mitigate fadeout in the case of a preschool math curriculum intervention—*Building Blocks*—in the TRIAD study. A key innovation of this study was that some students were assigned to an alternative "follow-through" treatment condition that included both the *Building Blocks* curriculum in preschool along with additional professional development for kindergarten and first grade teachers. This additional professional development was designed to help inform teachers of the mathematics content taught with *Building Blocks* during preschool, with the hope of reducing repetition. When compared with children who only received the preschool curricular intervention, students assigned to the follow-through condition had substantially less effect fadeout at the end of first grade. Previous analyses of the follow-through condition have not tested which specific elements of the classroom instructional environment explain these persistent treatment effects.

Although the TRIAD follow-through results contain potential promise, the bulk of the existing evidence points to a disconnection between children's knowledge and teachers' instructional content from preschool and kindergarten (Abry, Latham, Bassok, & LoCasale-Crouch, 2015). Kindergarten teachers spend considerable instructional time on content already mastered by preschool graduates, which may reduce learning among more advanced children (Engel et al., 2012; Engel, Claessens, Watts, & Farkas, 2016; Gervasoni & Perry, 2015; Magnuson, Ruhm, & Waldfogel, 2007). Early-grade curricula often assume students have limited prior knowledge and may not provide plans, methods or content which is designed to differentiate among students of differing skills levels. Thus, teachers may remain unaware that some of their students—preschool graduates—have already mastered the material they are required to teach (Sarama & Clements, 2015). Given the wide range of children's school-entry skills in a given classroom (Duncan et al., 2007), and increasing pressures to meet established proficiency benchmarks, a teacher may be forced to focus on minimal competency assessments, leaving preschool attendees without the appropriate challenge they need to maintain growth (D. H. Clements, Sarama, Wolfe, & Spitler, 2013; Sarama & Clements, 2015). In the preschool fadeout scenario, good instruction could involve the teacher going as fast or far as possible *as measured by children who are not yet meeting benchmarks*—in this case, those who did not attend preschool—and not covering more advanced content for the higher-skilled preschool participants. This good, but not differentiated, instruction would benefit the students with lower skills, facilitating catch-up or convergence with preschool attendees.

Indeed, this is what Magnuson, Ruhm, and Waldfogel (2007) found in their study of preschool fadeout using the Early Childhood Longitudinal Studies of Kindergarten (ECLS-K) 1998 cohort. Using class size and amount of time spent on academic instruction as proxies for classroom instructional quality, they tested whether these two features of kindergarten classrooms moderated the persistence of preschool gains. They found that preschool advantages persisted for children attending less enriching classes that were larger and had lower total instruction time. To explain this counterintuitive result, they hypothesized that children entering school with limited skills were benefitting more from small, academically-focused classes than their preschool-attending peers, allowing them to catch up; non-preschool attendees attending less academically-focused classes did not catch up, and thus the preschool advantage persisted.

Other correlational studies examining whether measures of early-grades instructional content and instructional enrichment or quality reduces preschool fadeout (i.e., differentially benefits preschool participants) lend little support for this idea. Claessens, Engel, and Curran (2013) also used the 1998 ECLS-K data and examined the relationship between the level (basic or advanced) and type (math or language and literacy) of content covered in kindergarten and the persistence of preschool effects. They found that "advanced" reading and math content (i.e., content that students did not encounter in preschool) in kindergarten was equally beneficial for all students, regardless of preschool attendance. A related study by Engels, Claessens, and Finch using the ECLS-K found that exposure to mathematics content already mastered during preschool impeded children's achievement growth (2012). Bassok et al. (2015) also examined moderation of preschool attendance by other proxies for kindergarten classroom enrichment using the ECLS-K, including both the 1998 and 2010 cohorts. They tested six features: full-day kindergarten, small class size, kindergarten school co-located with preschool, peer preschool attendance, use of kindergarten transition practices, time spent on reading in kindergarten. Like Claessens et al. (2013), they also found no meaningful differences in the rate of preschool fadeout based on children's subsequent kindergarten experiences.

In all of these studies, children were randomly assigned to neither their preschool (e.g., center-based care, Head Start, state pre-k, family child care) nor early-grade educational environments, so researchers relied on controls for observed characteristics of children to reduce possible biases of selection into both preschools and classrooms. This strategy, however, may be insufficient given the unobserved factors correlated with choosing to enroll in preschool, the heterogeneity in preschool experiences and the relative quality of such preschool, in addition to the variability in elementary school quality that families select into after preschool. Research that eliminates the biases from selection into preschool and, if possible, into elementary school is needed to understand whether kindergarten and first grade instruction might help to extend preschool gains.

A recent study by Bailey and colleagues (2016) addresses this possibility with the experimental *Building Blocks* TRIAD study data, testing what they refer to as the "constraining content" hypothesis. Nearly all children in the TRIAD study attended preschool at their district-assigned elementary school where they transitioned to kindergarten, substantially reducing the likelihood of school selection. They compared treatment and control group children who started kindergarten with the same level of math skills using a unique post-test matching design (omitting children assigned to the follow-through condition). They find that fadeout was more likely a result of pre-existing differences in children's skills rather than children's schooling experiences in kindergarten and first grade. The authors suggest that subsequent curricular *interventions*, not simply subsequent quality instruction, may be necessary to maintain the mathematics skills gains children made during preschool.

Present Study

In summary, some preschool interventions produce short-term achievement effects that fade over time, and theoretical explanations implicate low quality instruction in the early elementary years, but evidence using non-random assignment to preschool has not fully evaluated this hypothesis. A better way to think about the issue might be the ability of subsequent classrooms to further growth among preschool graduates who are relatively more skilled. Some research suggests that instructional content may be largely too basic and not sufficiently advanced or individualized for many children who have attended preschool. Furthermore, fadeout mechanisms may vary based on the characteristics of the intervention (e.g., preschool as usual versus a preschool curricular intervention). Understanding how these intervention characteristics relate to treatment effect fadeout is critical for developing early childhood programs that can produce sustained effects.

Using secondary data from the Head Start Impact Study (HSIS; Puma et al., 2010) and the scale-up of the *Building Blocks* preschool mathematics curriculum intervention called TRIAD (D. H. Clements & Sarama, 2008), we examined the extent to which the persistence of preschool program effects on children's cognitive skills depends upon the features of the kindergarten and first grade classrooms they attend. We used two key instructional characteristics, exposure to advanced language and literacy content (HSIS) and exposure to high-quality mathematics instruction (TRIAD), to operationalize sustaining elementary school learning environments in kindergarten and first grades. Using the HSIS, we first considered whether advanced literacy and language instruction in participants' kindergarten and first grade classrooms would better sustain end-of-treatment Head Start impacts. Specifically, we hypothesized that children who were assigned to attend Head Start and then experienced more rigorous instructional content would have higher early-grade achievement relative to both children assigned to attend Head Start but who then received relatively more basic early-grade instruction and children not assigned to attend Head Start.

We then turned to the evaluation of TRIAD, which randomly assigned schools with preschool classrooms to the *Building Blocks* mathematics curriculum or a control condition. The TRIAD study also randomly assigned half the schools in the treatment condition to additional professional support for mathematics instruction in kindergarten and first grade. Study evaluations show that classrooms with these follow-through professional supports produced more persistent program benefits on mathematics achievement for TRIAD participants compared with classrooms that did not have kindergarten and first-grade supports (D. H. Clements et al., 2013). In our study, we test whether observer-rated measures of mathematics instructional quality can explain the sustained treatment impacts observed for students in this extended treatment group.

The primary research questions for this study were:

1. Does the content level of academic instruction in kindergarten and first grade moderate the magnitude of Head Start intervention effects on children's language and literacy skills in kindergarten and first grade?
2. Do elementary school-level characteristics moderate the magnitude of Head Start intervention effects on children's language and literacy skills in kindergarten and first grade?
3. Does the quality of mathematics instruction in kindergarten and first grade moderate the magnitude of a preschool math curriculum intervention effect on children's math skills in kindergarten and first grade?
4. Does a professional development intervention for kindergarten and first grade teachers that provided techniques designed to build upon the preschool program moderate preschool curriculum intervention effects on children's math skills in kindergarten and first grade through mathematics instructional quality?

Only in the case of TRIAD were children randomly assigned to both preschool and early-grade curriculum enrichment. We investigate selection into subsequent school environments in our analyses. Although previous investigations of both the HSIS and TRIAD studies have examined the studies' initial treatment impacts and their subsequent fadeout patterns (Clements et al., 2011; 2013; Puma et al., 2010), our study is the first to investigate whether measures of later classroom instructional features (content and quality) moderate treatment impact fadeout. Using these two samples together also provides a useful replication exercise, examining whether long-standing hypotheses about preschool fadeout and persistence are consistent across two different early learning contexts.

Data

Preschool Intervention: Head Start

Head Start is a comprehensive child development program that provides children with preschool education, health screenings and examinations, and nutritious meals, in a full-day, center-based setting. The HSIS is representative sample of Head Start participants and a group of comparable non-participants from 23 states, sampled using a complex multi-stage stratified design as a part of the evaluation that began in 2002 (Puma et al., 2010). Head Start programs (grantees) that were oversubscribed (had waitlists) were divided into geographic clusters and were then stratified based on program characteristics, with three grantees or delegate agencies randomly selected from each cluster. Within each delegate agency, Head Start centers were stratified in the same way as grantees, and were randomly selected. This resulted in 84 programs and delegate agencies with a total of 383 individual preschool centers. The full sample included newly entering (i.e., no prior Head Start experience) 3- and 4-year-old Head Start applicants at randomly selected oversubscribed centers, where children were randomly assigned to receive an offer for Head Start. A total of 4,442 children were selected – 2,646 for Head Start and 1,796 for the control condition. Control group parents either made other ECE arrangements or cared for their children at home. Baseline survey and child assessment data were collected by study investigators in the Fall of 2002; post-treatment child assessments were collected at the end of Head Start in Spring 2003 and during kindergarten and first grade in Spring 2004 and 2005. Information on children's elementary school experiences was collected from kindergarten and first grade teachers through a teacher survey in the springs of 2004 and 2005.

The Head Start children in our sample first participated in the program during their pre-kindergarten year at age 4. Our analyses use the 4-year-old cohort only so that the children in both the HSIS and TRIAD analyses received the preschool intervention during the same developmental period and, in the case of the HSIS, had not been enrolled in Head Start in the year prior to study enrollment. The HSIS sample was further limited in our study to children whose elementary school teachers responded to the study survey and children who had not left the study at the kindergarten and first grade waves ($n = 1075$, 54% of the original 4-year-old cohort). Compared with the excluded sample, students in the analytic sample were 8% less likely to be black, 7% more likely to be white, 2% more likely to be DLL, and 4% more likely to have parents who are married. No other baseline characteristics were significantly different between the included and excluded sample.

The children and families were all low income (below the federal poverty level) and were of diverse racial and ethnic backgrounds (Table 1). Parents had low educational attainment with nearly 42% having less than a high school degree. About 23% of parents were recent immigrants and less than a fifth were teenage mothers, and a majority (84%) of the families lived in an urban area. Sixty-two percent of our analytic sample were in the treatment group, comparable to the original 4-year-old cohort (60%). The data in Table 1 show that treatment and control children in the analytic sample were very similar. The only significant difference between the two groups was in their preschool entry literacy & language skills composite score (.07 vs. −.05), which we control for in our analyses.

Children's language and literacy skills—Language and literacy skills were measured through direct on-on-one child assessments by study administrators in the child's main child care setting. The assessment battery involved a short series of tasks comprising the following instruments, widely used in child development research: the *Peabody Picture Vocabulary Test* (PPVT; Dunn & Dunn, 1997) and the *Letter Word* and *Spelling* subtests from the Woodcock-Johnson Psycho-Educational Battery-Revised III (WJ-LW and WJ-AP, respectively; Woodcock, McGrew, & Mather, 2001). The PPVT score was shortened from its original form, and the child's score was calculated using IRT methods and converted into a standard score. The WJ-LW and WJ-SP raw scores were converted to a linear IRT score as well as a standard score. Our analyses use the standard scores of each measure. To reduce chance findings owing to multiple testing, we created a language and literacy assessment composite measure to use as the main dependent variable by standardizing all three measures to mean 0 and standard deviation of 1, averaging across the three, and then re-standardizing the measure.[1]

Classroom environment—In the HSIS teacher survey, kindergarten teachers were asked how many times in the past week their class engaged in a given language or literacy activity. We coded each activity into basic or advanced for grade-level based on consultations with early literacy experts (faculty at the authors' institution; available in Appendix Table A). Because the teacher report responses to frequency of language or literacy activities in the HSIS were phrased in terms of times per week and times per month, we converted each basic and advanced activity responses to numeric values that represented times per month. For the responses that were "once or twice a week", "three or four times a week", and "every day", we took the mean weekly value of the answer category (e.g., never = 0; 1-2 times per week = 1.5), multiplied it by 4, and then standardized this measure to have a mean of 0 and standard deviation of 1, following Claessens, Engel, and Curran (2013). The responses "never", "once a month or less", and "two or three times a month", remained unchanged. Basic and advanced language and literacy content during the first grade is coded as a cumulative measure of content exposure from both kindergarten and first grade, averaging the measures across the two years (α=.82) to create a more comprehensive and stable measure of instruction based on both assessments of the instructional environment (i.e., two teacher surveys) in early grades.[2] Although the available measures of language and literacy instruction are relatively weak when compared with detailed studies of literacy instructional research (e.g., Connor et al., 2009), prior research indicates that teacher reports of classroom activities can be valid measures of the quantity of instruction (Herman, Klein, & Abedi, 2000).

Child and family covariates—Our analyses include child gender, race and ethnicity, preschool entry literacy skills assessment score (at study baseline), as well as limited English proficiency and special needs status. Family covariates include mother's race, ethnicity, age, education level (categorized as below a high school degree, high school degree or equivalent, and more than a high school degree or equivalent), marital status (married=1), and urbanicity (urban=1).

School and additional classroom moderating variables—Also included in the HSIS are non-instructional measures of early school experiences that may influence children's learning and persistence of preschool impacts. Shown in the bottom panel of Table 1, we incorporate the following variables in additional moderation analyses for the HSIS only: attending full-day kindergarten, kindergarten class size, classroom-level

[1]We also estimate our analyses using each individual outcome variable comprising the language and literacy composite separately (see robustness section below).

[2]We also ran our analyses with only the first-grade instructional measures (see robustness section below).

proportion of children in poverty (free or reduced-price lunch eligible), and school-level proportions of children in poverty and children proficient in reading and math.

Preschool Curricular Intervention: TRIAD

The TRIAD evaluation was designed to assess the long-term impacts of an instantiation of the TRIAD scale-up model, which involved implementing the *Building Blocks* preschool curriculum through extensive professional development and coaching (D. H. Clements, Sarama, Spitler, et al., 2011; Sarama, Clements, Wolfe, & Spitler, 2012). The study recruited 42 public elementary schools operating state preschool programs serving low-income communities in Massachusetts and New York in 2006. Schools were ranked according to state achievement test scores, and similarly-ranked schools were grouped into "blocks." This procedure, designed to ensure comparability between schools at study baseline, produced 8 blocking groups, and schools were randomly assigned within each block to one of three conditions: 1) *Building Blocks* preschool supplementary mathematics curriculum; 2) *Building Blocks* with follow-through; or 3) control (preschool as usual, including the districts' mathematics curricula). Children in schools assigned to the two *Building Blocks* groups received the curriculum during preschool, and preschool teachers attended 13 study-administered professional development (PD) sessions across two consecutive years. Children in control preschool classes received their usual math instruction, though the quality and content of this instruction varied (see Clements et al., 2011). Study participants across all three conditions were enrolled at the beginning of the preschool year at age-4 (n=1375).

Kindergarten and first grade teachers in schools assigned to "*Building Blocks* with follow-through" received PD designed to help bridge the gaps between preschool, kindergarten, and first grade. These PD sessions introduced teachers to what their children learned in the previous year(s) and to the *Building Blocks* learning trajectories, with the intent that they would use this information to alter their instruction and build on what students had already learned.[3]

Our analysis limited the TRIAD sample to children that had non-missing classroom observational measures (described below) in kindergarten or first grade and valid test score data in preschool, kindergarten and first grade (n= 821). Compared with the excluded sample, students in the analytic sample were 6% less likely to be male and 28% more likely to qualify for free or reduced-price lunch. Further, students in the analysis sample were also 7% more likely to have parents who were either at or below a high school education. No other baseline characteristics were significantly different between the included and excluded sample.

Table 2 presents descriptive statistics for the analytic sample by treatment condition, as well as *p*-values indicating whether baseline characteristics differed based on group assignment. We found no statistically significant differences among the groups. Across the three conditions, 56% identified as African American, 21% as Hispanic, 80% qualified for free or

[3]Clements and colleagues (2013) found that teachers were somewhat resistant to this additional PD, because they were teaching a new curriculum for the first time and believed that this already constituted a challenge and that simultaneously modifying it would be too challenging. Thus, they found that this follow-through treatment condition was weakly implemented.

reduced-price lunch, and 41% of children's mothers reported that they completed high school and had some college education.

Children's mathematics skills

Math achievement was assessed at preschool entry, and at the end of the preschool, kindergarten, and first grade years using the *Research Based Early Mathematics Assessment* (REMA; D. H. Clements, Sarama, & Liu, 2008; D. H. Clements, Sarama, & Wolfe, 2011). The REMA is designed to measure the mathematics achievement of children from ages 3 to 8 and is administered through two structured interviews that assesses competency in counting, operations, measurement, and geometry, among other topics. Interviews are videotaped and subsequently coded for both correctness and strategy use. The codes are then converted to a Rasch-IRT scaled score. The assessment was extensively validated in multiple samples, and has been shown to have a high correlation (.89) with the *Applied Problems* subtest of the Woodcock Johnson. The REMA has strong internal reliability (α= .94; D. H. Clements, Sarama, & Wolfe, 2011).

Classroom environment

Teachers' mathematics instructional practices in preschool, kindergarten, and first grade were evaluated via the *Classroom Observation of Early Mathematics Environment and Teaching* (COEMET; see D. H. Clements, Sarama, Spitler, et al., 2011). The COEMET is composed of 28 Likert-scaled items, and assessors, blind to treatment status, observed kindergarten and first grade classes once during each respective school year. Nine items of the 28 items, which measure the overall classroom culture, are rated once for every classroom per observation. The remaining 19 items are assessed every time an observer sees the teacher lead the class in a new math activity. These 19 items are only scored if a math activity is "substantial," defined as "one conducted intentionally by the teacher involving several interactions one or more students or set up conducted intentionally to develop mathematics knowledge" (p. 882, Clements et al., 2013). Together, the 19 items focus on teaching practices known to support early math development, such as the use of engaging small group activities and emphasizing cognitively demanding concepts and strategies. Because our analysis is focused on specific features of mathematics instruction in kindergarten and first grade following the *Building Blocks* program in preschool, we took the average of these 19 math instruction-focused items across every activity observed, and then standardized the scores. This approach was also used by Clements and colleagues (2011; 2013) in their analysis of the COEMET measure. Our measure of "math teaching quality" (i.e., the average of 19 instruction-focused COEMET items) had strong reliability in both kindergarten ($\alpha = 0.93$) and first grade ($\alpha = 0.88$).

Because these 19 items were assessed every time the class began a new, "substantial" math activity, we also included the number of math activities observed to measure the amount of math content the students received during kindergarten and first grade (also see Clements et al., 2011; 2013).

As with the HSIS, our measure of cumulative kindergarten and first grade instruction is the standardized average of a child's kindergarten and first grade "math teaching quality" scores, and we also took the average of the number of math activities observed across kindergarten and first grade. In Table 2, we present descriptive statistics for these two measures. Note that to be in our analysis sample, every kindergarten and first grade class needed to be observed conducting at least one math activity. In kindergarten, this restriction eliminated 13 classrooms containing 55 study students. In first grade, every class in the study had at least 1 math activity during the observational period. As shown in Table 2, we did not find significant differences between the three conditions on either the math teaching quality measure or the number of math activities recorded across kindergarten and first grade.

Child and family covariates

As shown in Table 2, covariates included measures of child gender, ethnicity, age and special education status at preschool entry, free or reduced price-lunch status, baseline mathematics score, whether designated limited English proficient, and mother's education.

Analysis

Our research questions focus on whether subsequent instructional experiences moderate the magnitude of preschool treatment effects in children's kindergarten and first grade year. This requires an approach that models both the programmatic preschool impact and an interaction between this programmatic impact and subsequent classroom characteristics. We use multivariate regression with interaction terms to test for this moderation in both studies.[4]

Note that the scope of our outcome analyses are limited by the measures of subsequent instructional experiences in each study. Our measures of sustaining classroom instruction in the HSIS are the frequency of exposure to advanced and basic language and literacy activities. Therefore, they capture *content level* but do not capture the quality of teachers' pedagogical strategies to support language and literacy skill development. Conversely, the measures available in the TRIAD study assess the quality of mathematics instruction, but not whether the classroom content was advanced or basic for grade level. The second instructional measure in TRIAD, number of math activities, measures quantity of instruction, and not content difficulty. Furthermore, neither study assessed whether instruction was individualized or differentiated based on children's skill level, which, as the literature suggests, is likely the most promising strategy for continuing learning gains.[5] These aspects of the studies' designs are thus a weakness of our analyses because they limit the systematic investigation of all potential classroom instructional factors that mitigate fadeout.

[4]We estimate intent-to-treat effects for each study because they are most policy relevant, as it tests whether the opportunity to participate in a given program at the population-level produces long-run impacts. Furthermore, we lack sufficient measures of program attendance in the TRIAD study for which to calculate treatment-on-treated (TOT) estimates.

[5]The COEMET does have one item that directly measures differentiation (*"The teacher adapted tasks and discussions to accommodate the range of children's abilities and development"*). Because it was only a single item, we opted to use the COEMET as it was intended, aggregating scores across items for each math activity.

Testing for Selection into Elementary School Environments

When testing for moderation, a key assumption for causal inference is that the moderating variable – in our case, classroom instruction – is as good as randomly assigned. Because preschool interventions were randomly assigned in our two secondary datasets, regression provides unbiased estimates of the offer of preschool attendance on children's outcomes (i.e., ITT), and in the TRIAD study, estimates of the exposure to an enhanced math curriculum. A key methodological issue, however, is that only the preschool intervention is randomly assigned; children's later classroom experiences are not and children and families may select into different environments or be systematically sorted into particular types of classroom environments post-treatment. To explore the potential for such bias in our moderators, we tested for selection into subsequent environments in both the HSIS (Appendix Table B) and the TRIAD study (Appendix Table C) by regressing the available kindergarten classroom and school characteristics (e.g., class size, school reading proficiency level) on children's treatment status. These tests revealed no evidence of differential selection into classroom or school environments by preschool treatment status.[6] Nevertheless, if children with better potential outcomes were more likely to experience more enriched classroom instructional environments in kindergarten and first grade, our estimates are likely to be upwardly biased and represent an upper bound of the true effect of the environmental moderation. Importantly, though, our data and analyses improve upon those of prior studies by removing bias from selection into preschool environments.

Note also that the TRIAD study data did not include the large set of classroom and school characteristics available in the HSIS, limiting our investigation of differential selection with these data. However, both treatment and control group children in the TRIAD study had selected into public school preschool programs without knowledge that the preschool intervention would occur. In addition, TRIAD preschool programs were located in the local public schools where children's kindergarten programs were also located. These features of the TRIAD study likely reduced the opportunities for children to differentially sort into public school kindergarten and first-grade environments. Indeed, we found no evidence that children assigned to either treatment condition were more likely to remain in the same school than children in the control group.

Analyses Testing the Sustaining Environments Hypothesis

We focused on language and literacy outcomes as the dependent variable in the HSIS models and on mathematics in the TRIAD models because language and literacy was the developmental domain with the largest treatment impact in the HSIS, and mathematics was explicitly targeted by TRIAD. These skills were therefore most likely to persist into later school years, providing the most potential variation to detect treatment effect moderation in kindergarten and first grade. In all models, we regressed achievement measures on treatment status controlling for baseline assessment scores and a set of child and family control variables. We then added measures of classroom instruction as covariates to see how much of the treatment effect was explained by exposure to basic or advanced language and literacy

[6]The only significant coefficient is full-day pre-k (β=−.05), but because this represents one of 22 regressions, this could be significant due to chance alone.

activities in HSIS and to variations of quality and quantity of instruction in TRIAD. Finally, we added models in which treatment was interacted with classroom instruction. If exposure to these measures of instruction in kindergarten and first grade helps reduce fadeout, then these interactions should be positive and significant.

HSIS regression models—The specification for the HSIS analysis is as follows:

$$Y_{itcj} = \beta_0 + \beta_1 Treat_j + \beta_2 Advance_{itc} + \beta_3 Basic_{itc} + \beta_4 Treat_j * Advance_{itc} + \beta_5 Treat_j * Basic_{itc} + \beta_6 X_i + \alpha_j + e_{itcj}$$

where Y is a language and literacy or mathematics outcome composite score for child *i* taken at either end of preschool, kindergarten, or first grade (*t*), *Treat* is the Head Start random assignment treatment indicator (students assigned to the control condition serve as the omitted comparison group), *j* indexes units of randomization, which in the case of the HSIS is the center, and *c* references classroom. *Advanced* and *Basic* represent our focal instructional variables—the total times per month (in SD units) that a teacher reported either advanced or basic language and literacy activities in the kindergarten or first grade classroom—indexed by classroom (*c*). We then add interaction terms between *Advanced* and *Basic* and the treatment indicator. With *Advanced* and *Basic* constructed to have a zero mean, β_1 is an estimate of the mean ITT treatment effect. Interaction coefficients β_4 and β_5 constitute our key coefficients of interest. X_i is a set of baseline child-level control variables listed in Table 1, α_j is a set of fixed effects for the units of random assignment, and e_{itcj} represents the unaccounted for factors contributing to children's language and literacy development. Clustered standard errors are used to address non-independence of observations within centers and classrooms.

Additional school and classroom moderator analyses using the HSIS: One concern with our primary models is that they have overlooked some overall experience in the early school years that is especially important by focusing only on specific instruction. Put another way, perhaps longstanding explanations that focus on overall school quality are more likely to show an association with the persistence of program impacts. The HSIS was a comprehensive study, and the dataset includes other characteristics about the kindergarten classroom environment, such as class size and proportion of children in poverty. We used the additional measures of classroom and school environments shown in the bottom panel of Table 1 to test for alternative hypotheses from prior research about the suppression or maintenance of treatment effects in elementary school (Chetty et al., 2011; Magnuson et al., 2007; Nye, Hedges, & Konstantopoulos, 2000; Stipek, 2004). The specification of these models follows our basic HSIS regression model shown above, replacing the classroom instructional variables with those listed in Table 1 (e.g., school proportion of students proficient in math and reading), and then interacts the focal variable with treatment to test for preschool treatment moderation.

Kindergarten classroom fixed effect model: Although there was no evidence of differential selection into elementary school environments based on observed characteristics, one might be concerned that unobserved or unmeasured features of the kindergarten classroom may still bias our estimates of fadeout. One way to address this concern is to compare the outcomes of treatment and control children experiencing the same instructional environment in a classroom fixed effects model (i.e., within-classroom analysis). Because the HSIS sample included treatment and control children who attended the same kindergarten class, we were able to estimate a kindergarten classroom fixed effect model with the HSIS data. The specification is as follows:

$$Y_{itc} = \beta_0 + \beta_1 Treat_j + p\left(\mathrm{K\ Classroom_c}\right) + \beta_2 X_i + e_{itcj}$$

where p is a vector of indicators for each of the (c) kindergarten classrooms in the study. In this model, β_1 is the coefficient of interest because it captures the difference in outcomes between children who did and did not participate in Head Start who share the same kindergarten classroom and (sustaining) instructional environment.

Analytic weights for HSIS analyses: The HSIS investigators at Westat created longitudinal sampling weights and corresponding jackknife standard errors for each wave to address differences in family nonresponse, attrition, and for complex sampling, which were used in the HSIS evaluation report analyses. However, recent studies using the HSIS do not use these weights because they cannot be replicated by other analysts (Bitler, Hoynes, & Domina, 2014; Bloom & Weiland, 2015).

We followed the weighting strategy used in Bitler et al. (2014) in their analyses of the HSIS data and used inverse probability of treatment weights (IPTW) to address imbalances between treatment and control groups due to attrition, and adjust for complex sampling. These weights accomplish the same goal as the original sampling weights, but are replicable.[7] (Further detail available in Appendix D). Another benefit from using IPTW is that they also incorporate the control variables shown in Table 1 in our models, so we did not need to include control variables in the outcome model specifications. Note that these weights do not adjust for teacher nonresponse.

TRIAD regression models—The specification for the TRIAD analysis is as follows:

$$Y_{itcj} = \beta_0 + \beta_1 Treat_j + \beta_2 MathQual_{itc} + \beta_3 NumMathAct_{itc} + \beta_4 Treat_j * MathQual_{itc} + \beta_5 Treat_j * NumMathAct_{itc} + \beta_6 X_i + \alpha_j + e_{itcj}$$

where Y is a mathematics outcome score for child i taken at either end of preschool, kindergarten, or first grade (t), *Treat* is the treatment indicator (students assigned to the control pre-K condition serve as the omitted comparison group), j indexes units of randomization, which in the case of TRIAD is the block group, and c references classroom. *MathQual* (average of COEMET items measuring quality of math instruction) and *NumMathAct* (number of math activities) represent our key instructional variables in

[7]Furthermore, the standard errors for the HSIS longitudinal weighted models with corresponding jackknife standard error calculations cannot be estimated in our models with include fixed effects or indicators for center of random assignment. For comparison, we present the results from tests of differential attrition by treatment status using the IPT weights and using the HSIS provided sampling weights in Appendix Table D.

kindergarten and first grade. As with the HSIS analysis, we then add interaction terms between *MathQual* and *NumMathAct* and the treatment indicator, with interaction coefficients β_4 and β_5 constituting our key coefficients of interest. X_i is a set of child-level baseline control variables shown in Table 2, α_j is a set of fixed effects for block group and e_{itcj} represents the unaccounted for factors contributing to children's mathematics development. As in the case of HSIS, clustered standard errors are used to address non-independence of observations within centers and classrooms. For the follow-through condition analyses we use the same specifications, comparing students in the follow-through treatment arm with control group students.

Results

Preschool Intervention: Head Start

Kindergarten instruction—The descriptive statistics of our kindergarten instruction variables in Table 1 indicate that average basic and advanced activities were 81 and 77, respectively, with maximum values of 100 (for both) and minimum values of 19 and 11, respectively. This suggests that there exists substantial variation in the number of activities observed in these data, and that average values are fairly high. To provide some context for interpretation, we compare the mean levels of kindergarten language and literacy instruction in the HSIS to the nationally representative calculations of classroom language and literacy activities from Claessens et al. (2013) as a benchmark. In Claessens et al., the ratio of basic to advanced activities—in days per month—was 18:11 (1.63); in the HSIS, the ratio—in times per month—was 81:77 (1.05). This comparison suggests that the kindergarten teachers in the HSIS sample report using similar amounts of basic and advanced classroom activities, whereas in a national sample of kindergarten classrooms, teachers report using a larger proportion of classroom instructional time devoted to basic skills.

Consistent with prior analyses of the HSIS data, our results show a modest effect of the Head Start offer on children's early literacy and language skills at the end of the preschool year (.16 SD, Model 1, Table 3a). However, by the end of kindergarten Head Start negatively predicts children's skills (Model 2).[8] When we add the elementary school instructional content variables (basic and advanced language and literacy instruction) in Model 3, we find that, as expected, more frequent use of advanced classroom literacy activities is associated with improvements in child skills, with an effect size of .09 (significant at the .10 level). In terms of our measure, this implies that a teacher-reported increase of advanced language and literacy activities of 17 times per month would be associated with a one-tenth of a SD increase in children's language and literacy skills. Conversely, basic language and literacy activities are negatively associated with children's skills (significant at the .05 level), with an interpretation and effect size of the same magnitude as advanced language and literacy instruction. Key to our analysis are the interactions between instructional content and Head Start offer. Neither treatment interaction with exposure to basic or advanced language and

[8]This corresponds with Puma et al.'s (2013) findings, reporting negative impact estimates for 9 of the 11 language and literacy outcomes at the end of third grade. However, their estimates were much smaller (−.01), and most of the HSIS first and third grade impacts are positive and insignificant. Taken together, we focus not on the −.15 impact coefficient and instead on the hypothesized interactions.

literacy content predicts children's language and literacy outcomes. Thus, the emphasis on either basic or advanced language and literacy strategies does not appear to affect the persistence of Head Start offer impacts by the end of kindergarten.

We also examined whether the relationships between advanced and basic instruction and children's literacy skill development were nonlinear by constructing "high" and "low" indicators for each measure based on a median split in the distribution for the measure (not shown). No significant relationships emerged when using these forms of the measure. In sum, greater exposure to more advanced language and literacy instruction did not sustain the gains of Head Start treatment group children through the kindergarten year, nor did an emphasis on basic skills erode the Head Start gains.

Classroom fixed effect—The kindergarten classroom fixed effect model (Model 5) tests whether Head Start participants have stronger language and literacy skills at the end of kindergarten relative to a control child in the same classroom (~250 children shared a classroom with a control child). The coefficient of interest is the treatment indicator, which denotes the difference in language and literacy skills between a treatment and control child who share the same classroom. This estimate therefore comprehensively controls for selection into elementary school and overall classroom experience, including sustaining environmental factors such as instruction, teacher's interactions, peer ability, class size, and any other unobserved features about the classroom (weights control for child and family characteristics). The treatment coefficient was not significant, meaning that robustly controlling for classroom and school characteristics, treatment children did not have a skill advantage at the end of kindergarten compared with control children not assigned to Head Start in the same classroom.

To test whether Head Start impacts are sustained in classrooms with more advanced instruction using fixed effects, we also estimated this model conditioning on classrooms that reported frequent use of advanced content in the analytic sample. We defined advanced content classrooms as those where teachers reported use of advanced language and literacy activities that was greater than one SD above the mean. The treatment coefficient was small, negative, and not significant.

First grade instruction—Results presented in Table 3b show the effect of the Head Start offer on the language and literacy composite at the end of first grade. As expected, we do not find that Head Start predicted children's level of literacy and language skills. However, the coefficients on advanced and basic literacy activities were very similar to those in the kindergarten models, with a positive main effect of advanced and negative main effect of basic language and literacy activities, significant at the .10 level. Still, we found no interaction between Head Start status and basic or advanced language and literacy activities indicating that more advanced content instruction, as measured by teacher report, did not sustain Head Start treatment effects through the end of first grade.

Additional classroom and school-level moderators—Table 4 presents the results for similar models that tested for moderation by other measures of early school experiences. When interacted with treatment, none of these measures was a significant predictor of kindergarten or first grade outcomes. Nor was there a coherent or consistent pattern of associations, suggesting that Head Start and control group children fared the same under these varying elementary school conditions.

Preschool Curricular Intervention: TRIAD

Kindergarten instruction—Overall, we found that in kindergarten, the average number of math activities observed was 2.48 (SD= 1.58; ranging from 1 to 7), and the average level of instructional quality (ranging from 1 to 5, with "5" indicating high quality) was 3.83 (SD=0.43). This appears to be a slightly high amount of exposure to math instruction when compared to what has been reported in nationally representative samples. Using the ECLS-K, Engel and colleagues (2012) reported that kindergarten teachers spent only 3 hours per week on math, and they covered between 1 and 2 math topics per day (using teacher self-reports rather than observational instruments as used in TRIAD).

Table 5a presents the impacts of TRIAD on kindergarten mathematics achievement. Model 1 shows the *Building Blocks* preschool treatment effect at the end of the preschool year with an effect size of .67. At kindergarten (Model 2), the effect dropped to .37 and remained significant. The treatment effect remained unchanged when we added the instructional quality variables (COEMET measure of math teaching quality and number of math activities) in Model 3, and the the number of math activities was a significant predictor of children's math achievement (0.12 SD), but math instructional quality was not. The SD for the number of math activities was 1.5 in kindergarten, suggesting that adding an additional 1.5 math activities per day would increase treated students' math achievement by about one-tenth of a SD by the end of kindergarten.

We did not find an interaction between the number of math activities and treatment status in model 4, but we found a marginally-statistically significant and positive interaction for the quality of math instruction variable with treatment status (0.12 SD). This suggests that improving math instruction by approximately half a point on the 5-point Likert scale (as rated by the COEMET) would improve treated students' math achievement by about one-tenth of a SD. This gives some indication that high-quality instruction may have provided a slight boost to children who received *Building Blocks* during preschool. However, a joint-test evaluating whether both interactions jointly contribute to the model was not statistically significant ($F(2, 40)= 0.66$, $p = 0.520$).

Models 5 and 6 examined the extent to which the kindergarten follow-through treatment with teacher PD including aspects that could reduce instructional repetition predicted math achievement. Unlike our measures of classroom instructional quality, teachers were randomly assigned to engage in additional PD. The treatment effect for students in the follow-through group was .40 and significant, but it was not significantly different from the end of kindergarten impact for students in the *Building Blocks*-only treatment group without follow-through (.58).

Does instructional quality account for differences in mathematics achievement among children assigned to the follow-through condition at the end of preschool? The main effect of the number of mathematics activities again was positive and significant (Model 7; .12), but not math instructional quality. Interactions between follow-through treatment with instructional quality (Model 8) were not significant, suggesting that instructional quality is not related to impact persistence at kindergarten.

We also examined whether the relationships between quality math instruction and children's math skills was nonlinear by splitting math instructional quality and number of math activities at the median, creating high and low categories, revealing no significant treatment interactions.

First grade instruction—Overall, TRIAD preschool treatment effects are smaller in first grade than in kindergarten. Instructional quality did not predict math scores, but the number of math activities did. However, these measures do not moderate the TRIAD program impact; the interaction between instructional quality and treatment was not significant, nor was the interaction between number of math activities and treatment. Thus, our measures of quality instruction did not sustain preschool math skill advantages for children assigned to the treatment condition without follow-through.

As with the last two kindergarten models in Table 5a, Model 5 in Table 5b focused on the TRIAD follow-through condition. The follow-through treatment effect size was .32 and significant, compared with .19 for *Building Blocks*-only. The .32 SD effect was only slightly smaller than the .38 follow-through effect found at the end of kindergarten, suggesting very little fadeout during first grade. However, comparing the follow-through and *Building Blocks*-only group impacts at the end of first grade revealed that the two effect sizes were not statistically significantly different ($p = 0.14$).

We found a surprising negative interaction between math instruction quality and treatment (-0.13 SDs) for the follow-through group, significant at the .10 level (Model 7 of Table 5b). As with the marginally significant interaction found in our kindergarten models, the joint-test was not significant ($F(2, 40) = 1.73$, $p = 0.19$), indicating that the interactions do not jointly contribute to the model.

Additional moderation analyses—The TRIAD scale-up evaluation did not collect information regarding the proportion of students in the kindergarten class that qualified for FRPL or whether classes were full-day or half-day. However, class size was obtained from the classroom observation. We tested an additional model (compare with models 2-4, Table 4) that included a main effect for class size and an interaction between treatment and class size. This revealed a negative main effect for class size (−.15, $p < .05$), and a positive interaction for class size and treatment (.17, $p < .05$), indicating that the treatment may have defended children from the negative effect of larger classes.

Robustness

We conducted several additional tests of our analyses to examine the robustness of our results to different assumptions and data constraints.

Missing kindergarten and first grade classroom data—Changes in observation counts across models in Tables 3 and 5 reflect changes in teacher survey item non-response

(HSIS), or missing classroom observations (TRIAD). We present analyses of outcomes based on kindergarten teacher response status for the HSIS in Appendix Table E, and the treatment effect remain the same, indicating that the students whose teacher did not respond to the survey are not meaningfully different than those whose teachers did not respond.

Appendix Table F shows equivalent analyses for the TRIAD study for students in the treatment and treatment with follow-through conditions. These results indicate that treatment effects at the end of preschool were larger for students with observational data in kindergarten and first grade. Students without observational data also experienced more effect fadeout between preschool and kindergarten.

Individual language and literacy outcomes in the HSIS—Following the same specifications as those presented in Table 3a and 3b, Appendix Tables G.1 and G.2 disaggregate the language and literacy outcome assessments that comprise the composite measure: PPVT, WJ Letter-Word Identification and Spelling subtests. Shown in G.1, models 3 and 4, the negative association between basic reading activities and children's outcomes at the end of kindergarten appear to be concentrated in children's receptive vocabulary (PPVT). During first grade, the positive association between advanced language and literacy activities and children's outcomes is most influential on children's early writing skills, as assessed by the WJ-Spelling subtest (G.2, models 14, 15).

Separating first grade instructional quality—The results from models using the HSIS measure of first grade instructional content only (not the pooled kindergarten and first grade measure used in our main models) are presented in Appendix Table H. We use the same specification as our main models shown in Table 3b, but replace the pooled kindergarten-first grade measures of basic and advanced classroom language and literacy activities with measures for only first grade basic and advanced classroom instruction. These results are similar to those presented in Table 3b, though the coefficient for advanced literacy activities loses significance.

The results from models using a measure of first grade instruction only for the TRIAD study are presented in Appendix Table I. Recall that sample inclusion for TRIAD analyses was contingent upon having either a non-missing kindergarten or first grade observation (n=821). For first grade, only 649 children had valid observations. We found a slightly larger end-of-preschool treatment effect for this group (.70), and a larger follow-through treatment effect in first grade for this sample (.46). We did not find that the first-grade classroom quality accounted for the follow-through effect, and we also found no positive interactions between treatment status and classroom quality measures.

Discussion

Our study tested whether instructional features of children's kindergarten and first grade classrooms could explain variation in program fadeout following two different preschool interventions; preschool as usual and preschool with a mathematics curricular intervention. As previously reported, we found substantial treatment impact fadeout in both samples (D. H. Clements et al., 2013; Puma et al., 2012). Head Start treatment effects were gone by kindergarten, and TRIAD treatment effects were reduced by 70% between preschool and the end of first grade. An important question is whether instruction played a role in explaining this pattern of declining impacts. Our measures of the classroom instructional environment were predictive of kindergarten and first grade achievement, where advanced-content instruction supported language and literacy skills and basic-content instruction inhibited them in the HSIS, as also found by Claessens et al. (2013). Similarly in TRIAD, the number of mathematics activities observed predicted mathematics skills in kindergarten and first grade. However, these measures were largely non-significant when interacted with treatment status, indicating that early grades instructional enrichment did not differentially benefit preschool participants. Although we found a marginally significant positive interaction with instructional quality in kindergarten, and a marginally significant negative interaction with instructional quality for the follow-through group in first grade, these coefficients were relatively small and were not observed consistently across all models. Taken together, instructional measures did little to explain fadeout.

In contrast, we found that the additional PD offered to teachers in the follow-through condition of TRIAD did help to sustain effects into first grade (see also Clements et al., 2013), though the initial treatment impact for this group still faded by approximately 50 percent. These results suggest that targeted PD—designed to create continuity and avoid repetition between grades—may be an effective way to sustain the impacts from high-quality preschool curriculum interventions. Unfortunately, we could not reveal the specific teaching processes that helped to sustain treatment impacts because our instructional quality measures did not account for follow-through treatment effect persistence. Nevertheless, the presence of the sustained follow-through treatment effect still suggests that certain instructional approaches may help sustain learning.

If alignment between preschool curricula and early-grade content is key, such alignment may be much more difficult to achieve with certain early childhood programs. In particular, Head Start centers usually operate independently of K-12 schools, so achieving coherence between Head Start and local kindergarten content would like be much more difficult task than aligning content between kindergarten and state pre-k classes housed in the same school. That difference is probably reflected in the data presented here, as the HSIS centers were not linked to specific schools, but the TRIAD preschool classes were in public elementary schools. Future research should examine the specific components of preschool and early grade alignment that produce sustained learning gains, and research should also focus on the practical challenges to achieving such alignment in diverse settings.

What might account for our primary findings that enriched, early-grades instruction did not moderate the persistence of preschool treatment effects? First, our measures of the classroom instructional environment may have been unable to capture the classroom experiences essential for sustaining early academic gains. Our classroom environment measures in the HSIS merely represent whether teachers reported spending time on certain types of activities. The TRIAD study included only a single observation in the kindergarten year and another during first grade (on fewer than half of the total number of TRIAD preschool classrooms), and these observations were solely focused on the quality of mathematics instruction, and not the specific content taught or broader features of the classroom environments and teaching. We noted that the differentiation of instruction is a key component of high-quality instruction—and is arguably the sine qua non-ingredient in maintaining early treatment effects, but could not measure whether teachers matched classroom instruction to children's skill levels because both studies only included classroom-level aggregate instructional measures. In this way, our measures do not capture true pedagogical differentiation, where teachers would recognize the more advanced skills of preschool graduates and subsequently increase their quantity of math activities (à la TRIAD), or the amount of time on advanced topics (à la HSIS).

Another common explanation for the lack of sustained impacts in the HSIS is that the control group attended a wide variety different alternative child care arrangements, including center-based child care, family child care home, parental care, or relative care. Several recent studies have examined how the estimated end-of-treatment impacts of Head Start vary depending on the comparison group using the HSIS, and find indeed that the main end of treatment effect for Head Start is strongest when compared with children in the control group who attended home-based care, with few to no differences compared with center-based care (Feller, Grindal, Miratrix, & Page, 2016; Kline & Walters, 2016; Walters, 2015; Zhai, Brooks-Gunn, & Waldfogel, 2014). Bloom and Weiland (2015) find substantial heterogeneity in Head Start treatment impacts by program site, with centers ranging from much more to much less effective than their local alternatives, including parent care. If control group children consistently attended higher-quality alternatives, our results would be biased towards zero. However, the research on HSIS counterfactual conditions does not suggest this is the case.

Another possibility for why our study did not find that classroom instructional mechanisms moderate fadeout may be that theories of preschool fadeout—essentially, the sustaining environments hypothesis—is wrong. Our study does not provide definitive evidence to debunk this prevailing theory; without formative assessment information, our study is a test of only some variables, and not of important instructional factors such as differentiation. Yet the failure of prior studies and our study to generate consistent evidence of this hypothesis raises the possibility that conventional models of fadeout do not accurately explain children's post-preschool learning trajectories. Still, the TRIAD follow-through condition does stand as an important piece of evidence that some instructional changes may be able to keep children on a higher achievement trajectory after preschool.

Further, moderation analyses require a substantial amount of power due to the fact that interactions effectively split the sample along dimensions of the moderator in question. This may have been an issue in the TRIAD analyses since the SE's for our interaction terms were typically around one-tenth of a SD. This means that to detect an effect at the.05 significance level, we would have needed interaction effects of at least one-fifth of a SD. It is possible that such effects in this context could be much smaller, and our study was simply insufficiently powered to detect effects smaller than 0.20.

In addition to those mentioned above, other limitations of our study relate to measurement. Because both of our samples are low income (below the poverty level) and attend schools with limited resources, we may not have as much variation in instructional quality as we might in a sample representative of all public-school kindergarten classrooms. This may have limited our ability to detect any persistence from better classroom instruction (measures standardized within each sample), and therefore may not provide population-representative variation in instructional quality. Nevertheless, we note that some measures of classroom instruction (advanced language and literacy instruction, total number of math activities) produced the expected positive main effects. This suggests that the measures did capture some meaningful variation in classroom instruction that related to student achievement in predictable ways. Furthermore, our measure of language and literacy instructional quality is poor relative to the scientific literature on reading instruction. Here, researchers measure several more detailed dimensions of reading, such as code focused versus meaning-based, or teacher- versus child-managed (Connor, Morrison, & Katch, 2004), which was beyond the detail of our data. The results of experiments conducted by Connor and colleagues demonstrate the benefit of interventions in literacy instruction that explicitly differentiates classroom instruction and in-class group work by a child's literacy skills (2009). Their work suggests that instruction may be most beneficial when it is tailored to the skill level of preschool graduates. Based on our study and measures alone, we cannot be confident that other dimensions of instruction, such as differentiation, would not prevent fadeout. Future studies would benefit from better language and literacy instructional quality measurement by trained observers and from measures of differentiation in both subjects.

Conclusion

Through analyses with longitudinal evaluations of two enriched preschool interventions, we did not find evidence to support the hypothesis that more advanced content instruction or better instructional quality mitigates the fadeout of preschool treatment effects on children's academic skills during elementary school (i.e., sustaining environments). However, we did find that advanced instruction was associated with positive gains, while basic instruction was associated with relative losses, in children's language and literacy skills, confirming past findings that all elementary school children benefit from advanced content instruction, regardless of preschool history. We also found some evidence that the coupling of the TRIAD intervention with teacher professional supports in kindergarten and first grade all but eliminated the fadeout of effects on math achievement observed between kindergarten and first grade, but this was not consistently explained by our measure of subsequent mathematics instructional quality. Future research should investigate aligned preschool-elementary school curricular approaches and techniques to facilitated differentiated instruction to sustain the benefits of preschool programs for children from low-income families.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Abry T, Latham S, Bassok D, LoCasale-Crouch J. Preschool and kindergarten teachers' beliefs about early school competencies: Misalignment matters for kindergarten adjustment. Early Childhood Research Quarterly. 2015; 31(2):78–88.

Andrews RJ, Jargowsky P, Kuhne K. The effects of Texas's targeted pre-kindergarten program on academic performance. 2012; doi: 10.3386/w18598National Bureau of Economic Research Working Paper Series, No 18598

Aos S, Miller M, Drake E. Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates. Federal Sentencing Reporter. 2006; 19(4):275.

Bailey D, Duncan GJ, Odgers CL, Yu W. Persistence and fadeout in the impacts of child and adolescent interventions. Journal of Research on Educational Effectiveness. 2017; 10(1):7–39. DOI: 10.1080/19345747.2016.1232459 [PubMed: 29371909]

Bailey D, Nguyen T, Jenkins JM, Domina T, Clements DH, Sarama JS. Fadeout in an early mathematics intervention: Constraining content or preexisting differences? Developmental Psychology. 2016; 52(9):1457–1469. DOI: 10.1037/dev0000188 [PubMed: 27505700]

Barnett WS. Long-term effects of early childhood programs on cognitive and school outcomes. Future of Children. 1995; 5:25–50.

Barnett WS. Effectiveness of early educational intervention. Science. 2011; 333(6045):975–978. DOI: 10.1126/science.1204534 [PubMed: 21852490]

Barnett WS, Masse LN. Comparative benefit-cost analysis of the Abecedarian program and its policy implications. The Economics of Early Childhood Education. 2007; 26(1):113–125.

Bassok, D., Gibbs, C., Latham, S. Do the benefits of early childhood interventions systematically fade? Exploring variation in the association between preschool participation and early school outcomesCharlottesville, VA: 2015EdPolicyWorks Working Paper Series

Belfield CR, Nores M, Barnett WS, Schweinhart LJ. The High/Scope Perry Preschool Program. Journal of Human Resources. 2006; 41(1):162–190. DOI: 10.3368/jhr.XLI.1.162

Bierman KL, Domitrovich CE, Nix RL, Gest SD, Welsh JA, Greenberg MT, Gill S. Promoting academic and social-emotional school readiness: The Head Start REDI Program. Child Development. 2008; 79(6):1802–1817. DOI: 10.1111/j.1467-8624.2008.01227.x [PubMed: 19037951]

Bitler MP, Hoynes HW, Domina T. Experimental evidence on distributional effects of Head Start. 2014; doi: 10.3386/w20434National Bureau of Economic Research Working Paper Series, No. 20434

Bloom, HS., Weiland, C. Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact StudyMDRC; New York, NY: 2015

Bronfenbrenner, U. Ecological systems theory. In: Vasta, R., editor. Annals of child developmentVol. 6. Greenwich, CT: JAI Press; 1989187-249

Brooks-Gunn J, Markman-Pithers L, Rouse CE. Starting Early: Introducing the Issue. The Future of Children. 2016; 26(2):3–19.

Camilli G, Vargas S, Ryan S, Barnett WS. Meta-analysis of the effects of early education interventions on cognitive and social development. Teachers College Record. 2010; 112(3):579–620.

Campbell FA, Conti G, Heckman JJ, Moon SH, Pinto R, Pungello E, Pan Y. Early childhood investments substantially boost adult health. Science. 2014; 343(6178):1478–1485. DOI: 10.1126/science.1248429 [PubMed: 24675955]

Campbell FA, Pungello E, Burchinal MR, Kainz K, Pan Y, Wasik BH, Ramey CT. Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up. Developmental Psychology. 2012; 48(4):1033. [PubMed: 22250997]

Campbell FA, Ramey CT, Pungello E, Sparling J, Miller-Johnson S. Early childhood education: Young adult outcomes from the Abecedarian Project. Applied Developmental Science. 2002; 6:42–57.

Campbell FA, Wasik BH, Pungello E, Burchinal MR, Barbarin OA, Kainz K, Ramey CT. Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. Early Childhood Research Quarterly. 2008; 23(4):452–466. DOI: 10.1016/j.ecresq.2008.03.003

Cascio EU, Schanzenbach DW. The impacts of expanding access to high-quality preschool education. Brookings Papers on Economic Activity. 2013; 2013:127–178.

Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D. How does your kindergarten classroom affect your earnings? Evidence from Project Star. The Quarterly Journal of Economics. 2011; 126(4):1593–1660. DOI: 10.1093/qje/qjr041 [PubMed: 22256342]

Claessens A, Engel M, Curran FC. Academic content, student learning, and the persistence of preschool effects. American Educational Research Journal. 2013; 51(2):403–434. DOI: 10.3102/0002831213513634

Clements DH, Sarama J. Experimental evaluation of the effects of a research-based preschool mathematics curriculum. American Educational Research Journal. 2008; 45(2):443–494. DOI: 10.3102/0002831207312908

Clements DH, Sarama J, Liu XH. Development of a measure of early mathematics achievement using the Rasch model: the Research-Based Early Maths Assessment. Educational Psychology. 2008; 28(4):457–482.

Clements DH, Sarama J, Spitler ME, Lange AA, Wolfe CB. Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. Journal for Research in Mathematics Education. 2011; 42(2):127–166.

Clements, DH., Sarama, J., Wolfe, CB. TEAM--Tools for early assessment in mathematicsColumbus, OH: McGraw-Hill Education; 2011

Clements DH, Sarama J, Wolfe CB, Spitler ME. Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies persistence of effects in the third year. American Educational Research Journal. 2013; 50(4):812–850.

Clements MA, Reynolds AJ, Hickey E. Site-level predictors of children's school and social competence in the Chicago Child–Parent Centers. Early Childhood Research Quarterly. 2004; 19(2):273–296. doi: http://dx.doi.org/10.1016/j.ecresq.2004.04.005.

Connor CM, Morrison FJ, Katch LE. Beyond the Reading Wars: Exploring the effect of child-instruction interactions on growth in early reading. Scientific Studies of Reading. 2004; 8(4):305–336. DOI: 10.1207/s1532799xssr0804_1

Connor CM, Piasta SB, Fishman B, Glasney S, Schatschneider C, Crowe E, Morrison FJ. Individualizing student instruction precisely: Effects of child × instruction interactions on first graders' literacy development. Child Development. 2009; 80(1):77–100. DOI: 10.1111/j.1467-8624.2008.01247.x [PubMed: 19236394]

Crosnoe R, Cooper CE. Economically disadvantaged children's transitions into elementary school: Linking family processes, school contexts, and educational policy. American Educational Research Journal. 2010; 47(2):258–291. DOI: 10.3102/0002831209351564 [PubMed: 20711417]

Crosnoe R, Leventhal T, Wirth RJ, Pierce KM, Pianta RC, NICHD Early Child Care Research Network. Family socioeconomic status and consistent environmental stimulation in early childhood. Child Development. 2010; 81(3):972–987. DOI: 10.1111/j.1467-8624.2010.01446.x [PubMed: 20573117]

Cunha F, Heckman JJ, Schennach SM. Estimating the technology of cognitive and noncognitive skill formation. Econometrica. 2010; 78(3):883–931. [PubMed: 20563300]

Currie J. Early childhood education programs. Journal of Economic Perspectives. 2001; 15(2):213–238.

Currie J, Thomas D. Does Head Start make a difference? The American Economic Review. 1995; 85(3):341–364. DOI: 10.2307/2118178

Currie J, Thomas D. School quality and the longer-term effects of Head Start. Journal of Human Resources. 2000; 35(4):755–774.

Darling-Hammond L. Standards, accountability, and school reform. Teachers College Record. 2004; 106(6):1047–1085.

Deming D. Early childhood entervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Economics. 2009; 1(3):111–134. DOI: 10.2307/25760174

Diamond A. The evidence base for improving school outcomes by addressing the whole child and by addressing skills and attitudes, not just content. Early Education and Development. 2010; 21(5): 780–793. DOI: 10.1080/10409289.2010.514522 [PubMed: 21274420]

Diamond A, Barnett WS, Thomas J, Munro S. Preschool program improves cognitive control. Science. 2007; 318(5855):1387–1388. DOI: 10.1126/science.1151148 [PubMed: 18048670]

Dodge KA, Bai Y, Ladd HF, Muschkin CG. Impact of North Carolina's Early Childhood Programs and Policies on Educational Outcomes in Elementary School. Child Development. 2017; 88(3):996–1014. DOI: 10.1111/cdev.12645 [PubMed: 27859011]

Duncan GJ, Magnuson K. Investing in preschool programs. The Journal of Economic Perspectives. 2013; 27(2):109–132. DOI: 10.1257/jep.27.2.109 [PubMed: 25663745]

Dunn, LM., Dunn, LM. Peabody Picture Vocabulary Test--Third Edition (PPVT-III)Upper Saddle River, NJ: Pearson Publishing; 1997

Early DM, Iruka IU, Ritchie S, Barbarin OA, Winn DMC, Crawford GM, Pianta R. How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. Early Childhood Research Quarterly. 2010; 25(2):177–193. DOI: 10.1016/j.ecresq.2009.10.003

Elango S, García JL, Heckman JJ, Hojman A. Early Childhood Education. 2015; doi: 10.3386/w21766National Bureau of Economic Research Working Paper Series, No. 21766

Elkind, D. The hurried child: Growing up too fast too soonCambridge, MA: Da Capo Press; 2007

Engel M, Claessens A, Finch MA. Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. Educational Evaluation and Policy Analysis. 2012; 35(2):157–178. DOI: 10.3102/0162373712461850

Engel M, Claessens A, Watts T, Farkas G. Mathematics Content Coverage and Student Learning in Kindergarten. Educational researcher. 2016; 45(5):293–300. DOI: 10.3102/0013189x16656841 [PubMed: 29353913]

Fantuzzo JW, Gadsden VL, McDermott PA. An integrated curriculum to improve mathematics, language, and literacy for Head Start children. American Educational Research Journal. 2011; 48(3):763–793. DOI: 10.3102/0002831210385446

Feller A, Grindal T, Miratrix L, Page L. Compared to what? Variation in the impacts of early childhood education by alternative care-type settings. Annals of Applied Statistics. 2016; 10(3):1245–1285.

Garces E, Thomas D, Currie J. Longer-term effects of Head Start. The American Economic Review. 2002; 92:999–1012.

Gervasoni, A., Perry, B. Children's mathematical knowledge prior to starting school and implications for transition. In: Perry, B.MacDonald, A., Gervasoni, A., editors. Mathematics and Transition to SchoolNew York, NY: Springer; 201547-64

Goffin, SG., Wilson, C. Curriculum models and early childhood education: Appraising the relationshipNew York, NY: Merrill; 1994

Heckman JJ. Skill Formation and the Economics of Investing in Disadvantaged Students. Science. 2006; 312(5782):1900–1902. [PubMed: 16809525]

Heckman JJ, Pinto R, Savelyev P. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. The American Economic Review. 2013; 103(6):1–35. [PubMed: 25506083]

Herman JL, Klein DCD, Abedi J. Assessing students' opportunity to learn: Teacher and student perspectives. Educational Measurement: Issues and Practice. 2000; 19(4):16–24. DOI: 10.1111/j.1745-3992.2000.tb00042.x

Hill CJ, Gormley WT, Adelstein S. Do the short-term effects of a high-quality preschool program persist? Early Childhood Research Quarterly. 2015; 32(3):60–79.

Jenkins JM. Early childhood development as economic development: Considerations for state-level policy innovation and experimentation. Economic Development Quarterly. 2014; 28(1):147–165.

Jenkins, JM., Duncan, GJ. Do pre-kindergarten curricula matter?. In: Phillips, D.Dodge, KA., Pre-Kindergarten Task Force. , editors. The Current State of Scientific Knowledge on Pre-Kindergarten EffectsWashington, D.C: Brookings Institution and Duke University; 201737-44

Johnson, RC., Jackson, K. Reducing inequality through dynamic complementarity: Evidence from Head Start and Public School SpendingCambridge, MA: 2017NBER Working Paper No. 23489

Kline P, Walters C. Evaluating public programs with close substitutes: The case of Head Start. Quarterly Journal of Economics. 2016; 131(4):1795–1848.

Ladd HF, Muschkin CG, Dodge KA. From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. Journal of Policy Analysis and Management. 2014; 33(1):162–187.

Lee VE, Loeb S. Where do Head Start attendees end up? One reason why preschool effects fade out. Educational Evaluation and Policy Analysis. 1995; 17(1):62–82. DOI: 10.3102/01623737017001062

Li, W., Leak, J., Duncan, GJ., Magnuson, K., Schindler, H., Yoshikawa, H. Working Paper National Forum on Early Childhood Policy and Programs, Meta-analytic Database ProjectCenter on the Developing Child, Harvard University; 2016Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program.

Lipsey, MW., Farran, DC., Hofer, KG. A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third gradeNashville, TN: 2015

Ludwig J, Phillips DA. Long-term effects of Head Start on low-income children. Annals of the New York Academy of Sciences, 1136(Reducing the Impact of Poverty on Health and Human Development: Scientific Approaches). 2008:257–268.

Magnuson KA, Ruhm C, Waldfogel J. The persistence of preschool effects: Do subsequent classroom experiences matter? Early Childhood Research Quarterly. 2007; 22(1):18–38. doi: http://dx.doi.org/10.1016/j.ecresq.2006.10.002.

McKey, RH., Condelli, L., Ganson, H., Barrett, B., McConkey, C., Plantz, MC. The impact of Head Start on children, families and communities Final report of the Head Start Evaluation, Synthesis and Utilization ProjectDepartment of Health and Human Services; 1985

McLoyd VC. Socioeconomic disadvantage and child development. American Psychologist. 1998; 53(2):185–204. [PubMed: 9491747]

Miller EB, Farkas G, Vandell DL, Duncan GJ. Do the effects of head start vary by parental preacademic stimulation? Child Development. 2014; 85(4):1385–1400. [PubMed: 24597729]

Morris, PA., Mattera, SK., Castells, N., Bangser, M., Bierman, KL., Raver, CC. Impact findings from the Head Start CARES Demonstration: National evauation of three approaches to improving preschoolers' social and emotional competenceWashington, DC: 2014OPRE Report 2014-44

Muschkin CG, Ladd HF, Dodge KA. Impact of North Carolina's early childhood initiatives on special education placements in third grade. Educational Evaluation and Policy Analysis. 2015; 37(4): 478–500.

National Mathematics Advisory Panel. Foundations for success: The final report of the National Mathematics Advisory PanelWashington, DC: 2008

Nelson AA, Gazley B. The Rise of school-supporting nonprofits. Education Finance and Policy. 2014; 9(4):541–566. DOI: 10.1162/EDFP_a_00146

Nye B, Hedges LV, Konstantopoulos S. The effects of small classes on academic achievement: The results of the Tennessee class size experiment. American Educational Research Journal. 2000; 37(1):123–151. DOI: 10.3102/00028312037001123

Pianta R, Belsky J, Houts R, Morrison F. Opportunities to learn in America's elementary classrooms. Science. 2007; 315(5820):1795–1796. DOI: 10.1126/science.1139719 [PubMed: 17395814]

Puma, M., Bell, S., Cook, R., Heid, C. Head Start Impact Study Final ReportU.S. Department of Health and Human Services, Administration for Children and Families; Washington, DC: 2010

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Downer, JT. Third grade follow-up to the Head Start Impact Study: Final ReportOffice of Planning, Research and Evauation, Administration for Children and Families; Washington, DC: 2012

Reynolds AJ, Ou S, Topitzes D. Path of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. Child Development. 2004; 75:1299–1328. [PubMed: 15369516]

Reynolds AJ, Temple JA, Ou S. Preschool education, educational attainment, and crime prevention: Contributions of cognitive and non-cognitive skills. Children and Youth Services Review. 2010; 32(8):1054–1063. DOI: 10.1016/j.childyouth.2009.10.019 [PubMed: 27667885]

Ritchie, S., Willer, B. Curriculum: A guide to the NAEYC early childhood program standard and related accreditation criteriaWashington, D.C: National Association for the Education of Young Children (NAEYC); 2008

Sarama, J., Clements, DH. Scaling up early mathematics interventions: Transitioning with trajectories and technologies. In: Perry, B.MacDonald, A., Gervasoni, A., editors. Mathematics and transition to schoolSingapore: Springer; 2015153-169

Sarama J, Clements DH, Wolfe CB, Spitler ME. Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. Journal of Research on Educational Effectiveness. 2012; 5(2):105–135. DOI: 10.1080/19345747.2011.627980

Schweinhart, LJ. Lifetime effects: The High/Scope Perry Preschool Study through age 40Ypsilanti, M.I: High/Scope Press; 2005

Stipek D. Teaching practices in kindergarten and first grade: Different strokes for different folks. Early Childhood Research Quarterly. 2004; 19(4):548–568.

Stipek D. No Child Left Behind Comes to Preschool. The Elementary School Journal. 2006; 106(5): 455–466. DOI: 10.1086/505440

Swain WA, Springer MG, Hofer KG. Early grade teacher effectiveness and pre-k effect persistence: Evidence from Tennessee. AERA Open. 2015; 1(4)doi: 10.1177/2332858415612751

Vygotsky LS. Interaction between learning and development. Readings on the development of children. 1978; 23(3):34–41.

Walters C. Inputs in the production of early childhood human capital: Evidence from Head Start. American Economic Journal: Applied Economics. 2015; 7(4):76–102. DOI: 10.3386/w20639

Woodcock, RW., McGrew, KS., Mather, N. Woodcock-Johnson-III Tests of AchievementItasca, IL: Riverside; 2001

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, MR., Espinosa, LM., Gormley, W., Zaslow, MJ. Investing in our future: The evidence base on preschool educationFoundation for Child Development, Society for Research in Child Development; New York, NY: 2013

Zhai F, Brooks-Gunn J, Waldfogel J. Head Start's impact is contingent on alternative type of care in comparison group. Developmental Psychology. 2014; 50(12):2572. [PubMed: 25329552]

Zhai F, Raver CC, Jones SM. Academic performance of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings. Children and Youth Services Review. 2012; 34(5):946–954. [PubMed: 22773872]

Zigler, EF., Bishop-Josef, SJ. The cognitive child versus the whole child: Lessons from 40 years of Head Start. In: Singer, DG.Golinkoff, R., Hirsh-Pasek, K., editors. Play= learning: How play motivates and enhances children's cognitive and social-emotional growthNew York, NY: Oxford University Press; 200615-35

Zigler, EF., Styfco, SJ., editors. The Head Start DebatesMaryland: Paul H. Brookes Publishing; 2004

Table 1

Descriptive Statistics for the Head Start Impact Study Analytic Sample

	Treatment				Control				
Child and Family Characteristics	**mean**	**SD**	**min**	**max**	**mean**	**SD**	**min**	**max**	**Mean diff. T/C, *p-val***
Male	0.52		0	1	0.52		0	1	0.95
White	0.39		0	1	0.41		0	1	0.68
African American	0.16		0	1	0.16		0	1	0.91
Hispanic	0.45		0	1	0.44		0	1	0.62
Preschool entry literacy & language skills composite (std.)	0.07	1.04	−2.83	4.24	−0.05	0.94	−3.10	2.90	**0.05**
Limited English proficiency	0.32		0	1	0.33		0	1	0.51
Child has special needs	0.14		0	1	0.13		0	1	0.24
Mother's Race									
White	0.41		0	1	0.41		0	1	1.00
African American	0.16		0	1	0.16		0	1	0.99
Hispanic	0.43		0	1	0.43		0	1	0.99
Mother's age	29.65	7.45	18.00	75.00	29.42	6.64	19.00	62.00	0.61
Mothers education									
Below High School	0.43		0	1	0.42		0	1	0.77
High School	0.30		0	1	0.34		0	1	0.19
> High School	0.28		0	1	0.25		0	1	0.30
Lives in urban area	0.85		0	1	0.85		0	1	0.88
Marital status	0.48		0	1	0.53		0	1	0.14
Classroom Instructional Characteristics									
Total basic language & literacy activities in K (times per month)	80.49	17.11	19	100	81.57	16.00	14	100	0.32
Total advanced language & literacy activities in K (times per month)	76.63	16.95	11	100	76.79	15.52	21	100	0.88
Total basic language & literacy activities in K & G1 (times per month)	179.28	70.07	19	314	181.22	68.84	14	314	0.66
Total advanced language & literacy activities in K & G1 (times per month)	182.97	76.48	11	340	186.78	74.67	21.5	334	0.43
School Characteristics									
Full-day Kindergarten	0.60		0	1	0.64		0	1	0.19
Kindergarten class size	20.73	5.24	2	50	20.66	4.94	1	50	0.84
Classroom % children eligible for FRPL	68.72	32.12	0	100	68.62	31.56	0	100	0.96

	Treatment				Control				
Child and Family Characteristics	**mean**	**SD**	**min**	**max**	**mean**	**SD**	**min**	**max**	**Mean diff. T/C, *p-val***
School % children eligible for FRPL	65.56	26.31	0	100	65.78	6.21	0	100	0.90
School % proficient in reading	57.67	28.13	2	100	59.66	28.02	4	100	0.31
School % proficient in math	65.10	22.03	0	100	66.07	22.61	0	100	0.53
Observations	653				390				

Note. SD=Standard deviation (shown for continuous variables only); FRPL=Free or reduced-price lunch; K=Kindergarten. The *p*-value column presents results from tests of differences in proportions and means between the treatment and control group.

Table 2

Descriptive Statistics for the TRIAD Study Analytic Sample

	Treatment				Treatment with Follow-Through				Control				
Child and Family Characteristics	**mean**	**SD**	**min**	**max**	**mean**	**SD**	**min**	**max**	**mean**	**SD**	**min**	**max**	**p-value**
Male	0.48		0	1	0.45		0	1	0.48	0.50		1	0.82
African American	0.58		0	1	0.57		0	1	0.53	0.50		1	0.92
Hispanic	0.16		0	1	0.22		0	1	0.24	0.43		1	0.64
Ethnicity- Other	0.03		0	1	0.08		0	1	0.09	0.28		1	0.22
Age (years) at baseline	4.32	0.34	3.67	5.83	4.30	0.34	3.67	5.50	4.40	0.36	3.75	6.67	0.28
Preschool entry math skills (std.)	−0.10	1.10	−4.78	2.26	−0.02	1.02	−4.78	3.05	0.05	0.91	−3.98	2.04	0.64
Free/reduced price-lunch recipient	0.81		0	1	0.75		0	1	0.84	0.37		1	0.37
Limited English proficiency	0.11		0	1	0.16		0	1	0.22	0.42		1	0.45
Child has special needs	0.16		0	1	0.18		0	1	0.13	0.33		1	0.19
Mother' s Education													
Below High School	0.13		0	1	0.13		0	1	0.11	0.31		1	0.69
High School	0.28		0	1	0.26		0	1	0.26	0.44		1	0.79
> High School	0.41		0	1	0.43		0	1	0.38	0.49		1	0.75
Classroom Characteristics													
COEMET- Math teaching quality (K)	3.80	0.33	2.79	4.41	3.90	0.47	2.11	4.93	3.77	0.46	2.25	4.38	0.55
Number of math activities (K)	2.48	1.59	1.00	7.00	2.58	1.36	1.00	6.00	2.35	1.80	1.00	8.00	0.87
COEMET- Math teaching quality (K & FG)	3.80	0.33	2.79	4.32	3.94	0.34	2.11	4.93	3.81	0.37	2.25	4.38	0.26
Number of math activities (K & FG)	2.35	1.35	1.00	7.00	2.15	0.85	1.00	5.00	2.16	1.20	1.00	8.00	0.86
Observations	263				304				254				

Note. The p-value column represents results from a series of regressions in which each respective baseline characteristic was regressed on dummy indicators for both treatment groups (control children were the omitted comparison group). We then calculated a joint test of statistical significance testing whether both treatment groups were jointly different from 0 on each observable characteristic.

Table 3

Head Start Impact Study Results

a. Sustained Classroom Environment - End of Kindergarten Language and Literacy Composite

	(1) End of HS	(2) Spring of K	(3) Spring of K	(4) Spring of K	(5) Adjusted for Kindergarten Classroom Fixed Effect
Treatment	0.16*	−0.15*	−0.15*	−0.15*	−0.10
	(0.06)	(0.06)	(0.06)	(0.06)	(0.27)
Total Advanced Literacy Activities in K (times per month; std.)			0.09+	0.06	
			(0.05)	(0.07)	
Total Basic Literacy Activities in K (times per month; std.)			−0.10*	−0.08	
			(0.05)	(0.07)	
Treat * Advanced Literacy Activities				0.04	
				(0.09)	
Treat * Basic Literacy Activities				−0.04	
				(0.10)	
Observations	1043	1043	1041	1041	1043

b. Sustained Classroom Environment - End of First Grade Language and Literacy Composite

	(1) End of HS	(2) Spring of 1st Grade	(3) Spring of 1st Grade	(4) Spring of 1st Grade
Treatment	0.16*	−0.08	−0.07	−0.07
	(0.08)	(0.06)	(0.06)	(0.07)
Total Advanced Literacy Activities in K and G1 combined (times per month; std.)			0.12+	0.12
			(0.07)	(0.11)
Total Basic Literacy Activities in K and G1 combined (times per month; std.)			−0.12+	−0.11
			(0.07)	(0.10)
Treat * Advanced Literacy Activities				−0.00
				(0.14)
Treat * Basic Literacy Activities				−0.02
				(0.12)

a. Sustained Classroom Environment - End of Kindergarten Language and Literacy Composite

	(1) End of HS	(2) Spring of K	(3) Spring of K	(4) Spring of K	(5) Adjusted for Kindergarten Classroom Fixed Effect
Observations	1043	1065	1065	1065	

Note: Models 1-4 only include children whose teacher responded to the survey. Analyses of outcomes based on teacher response status available in Appendix Table E. Analyses using each individual outcome variable comprising the literacy composite are available in Appendix Table G. Standard errors clustered at class level (in parentheses). All models include fixed effects for center of random assignment and are weighted using inverse probability of treatment weights to adjust for differential attrition and complex sampling. Outcome measures are in standard deviation units. Weights include all the baseline child and family control variables.

+ $p < .10$;

* $p < .05$;

** $p < .01$

Table 4

Moderation by Classroom and School Factors in the Head Start Impact Study: Kindergarten Language and Literacy Composite Scores

	(1) Full-Day Kindergarten	(2) Kindergarten class size	(3) Classroom-level poverty (% FRPL eligible)	(4) School-level poverty (% FRPL eligible)	(5) School % Students Proficient in Reading	(6) School % Students Proficient in Math
Treat	−0.21*	−0.34	−0.09	−0.27	−0.24	−0.30
	(0.10)	(0.29)	(0.19)	(0.20)	(0.15)	(0.20)
Full-day K	0.18					
	(0.13)					
Treat *Full-day K	0.21					
	(0.16)					
Class size		0.01				
		(0.01)				
Treat *Class size		0.01				
		(0.01)				
Classroom % FRPL			0.02			
			(0.25)			
Treat *Class % FRPL			−0.30			
			(0.26)			
School % FRPL				0.26		
				(0.29)		
Treat *School % FRPL				−0.50+		
				(0.29)		
School % reading proficient					0.00	
					(0.00)	
Treat *School % reading proficient					0.00	
					(0.00)	
School % math proficient						0.00
						(0.00)
Treat *School % math proficient						0.00
						(0.00)

	(1) Full-Day Kindergarten	(2) Kindergarten class size	(3) Classroom-level poverty (% FRPL eligible)	(4) School-level poverty (% FRPL eligible)	(5) School % Students Proficient in Reading	(6) School % Students Proficient in Math
Observations	1008	971	777	898	927	928

Note: Coefficients, and standard errors in parentheses, presents the results for models similar to those presented in Table 3a that test for moderating effects of the column name in the HSIS. Standard errors clustered at school level (in parentheses). All models include fixed effects for center of random assignment and are weighted using inverse probability of treatment weights to adjust for differential attrition and complex sampling. Weights include all the baseline child and family control variables. Changes in observation counts across models reflect changes in teacher survey item non-response. FRPL-Free and Reduced-price lunch.

+ $p < .10$;

* $p < .05$;

** $p < .01$.

Table 5

Building Blocks TRIAD Scale-Up Study Results

5a. Sustained Classroom Environment – End of Kindergarten Math Composite

	(1) End of Pre-k	(2) Spring of K	(3) Spring of K	(4) Spring of K	(5) End of Pre-k	(6) Spring of K	(7) Spring of K	(8) Spring of K
Treatment	0.67***	0.37***	0.34***	0.35***	0.67***	0.35***	0.33***	0.33***
	(0.08)	(0.09)	(0.08)	(0.08)	(0.08)	(0.09)	(0.07)	(0.08)
Mathematics Teaching Quality			0.04	0.00			0.03	0.05
			(0.04)	(0.04)			(0.03)	(0.03)
Number of Math Activities			0.12*	0.12			0.13**	0.12*
			(0.05)	(0.09)			(0.04)	(0.05)
Treat*Mathematics Teaching Quality				0.12+				
				(0.07)				
Treat*Number of Math Activities				–0.00				
				(0.09)				
Treatment with Follow-Through					0.64***	0.40***	0.38***	0.38***
					(0.09)	(0.09)	(0.07)	(0.07)
Follow-Through*Mathematics Teaching Quality								–0.06
								(0.05)
Follow-Through*Number of Math Activities								0.03
								(0.12)
Observations	517	517	517	517	821	821	821	821

5b. Sustained Classroom Environment – End of First Grade Math Composite

	(1) End of Pre-k	(2) Spring of 1st Grade	(3) Spring of 1st Grade	(4) Spring of 1st Grade	(5) Spring of 1st Grade	(6) Spring of 1st Grade	(7) Spring of 1st Grade
Treatment	0.67***	0.18*	0.15*	0.18*	0.19*	0.16*	0.16*
	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Mathematics Teaching Quality			0.03	–0.01		0.00	0.04
			(0.04)	(0.05)		(0.03)	(0.04)

5b. Sustained Classroom Environment – End of First Grade Math Composite

	(1) End of Pre-k	(2) Spring of 1st Grade	(3) Spring of 1st Grade	(4) Spring of 1st Grade	(5) Spring of 1st Grade	(6) Spring of 1st Grade	(7) Spring of 1st Grade
Number of Math Activities			0.14^{**}	0.19^{+}		0.14^{**}	0.13^{**}
			(0.05)	(0.10)		(0.05)	(0.05)
Treat * Mathematics Teaching Quality				0.10			
				(0.07)			
Treat * Number of Math Activities				–0.09			
				(0.13)			
Treatment with Follow-Through					0.32***	0.32***	0.33***
					(0.09)	(0.07)	(0.07)
Follow-Through * Mathematics Teaching Quality							-0.13^{+}
							(0.07)
Follow-Through * Number of Math Activities							0.06
							(0.13)
Observations	517	517	517	517	821	821	821

Note: Standard errors clustered at school level (in parentheses). Mathematics teaching quality and number of math activities were measured using the COEMET. For each variable, scores were averaged from classroom observations in the kindergarten year. All models include controls for gender, ethnicity, age at preschool entry, mother's education level, free or reduced price lunch status, special education status at preschool entry, whether limited English proficient, and indicators for blocking group at random assignment. The sample was restricted to students who had a non-missing classroom observation (COEMET) in kindergarten or first grade and who were non-missing on all preschool, kindergarten, and first grade mathematics measures. Analyses of outcomes based on classroom observation status available in Appendix F. Outcome measures are in standard deviation units.

$^{+}p < .10$;

$^{*}p < .05$;

$^{**}p < .01$.

Note: Standard errors clustered at school level (in parentheses). Mathematics teaching quality and number of math activities were measured using the COEMET. For each variable, scores were averaged from classroom observations in the kindergarten and first grade year. All models include controls for gender, ethnicity, age at preschool entry, mother's education level, free or reduced price lunch status, special education status at preschool entry, whether limited English proficient, and indicators for blocking group at random assignment. The sample was restricted to students who had a non-missing classroom observation (COEMET) in kindergarten or first grade and who were non-missing on all preschool, kindergarten, and first grade mathematics measures. Outcome measures are in standard deviation units.

$^{+}p < .10$;

$^{*}p < .05$;

** $p < .01$

Funding

This research was supported by awards from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) to Greg Duncan and George Farkas (P01HD065704) and Elizabeth Gershoff (R01HD06956). Research was also supported by the Institute for Education Sciences (IES) from awards to Duncan (R305B120013) and Douglas Clements (R305K05157 & R305A120813). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of NICHD or IES.